AD-A119 375 STANFORD UNIV CA DEPT OF STATISTICS F/G 12/1
SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL.(U)
JUL 82 T SELLKE, D SIEGMUND N00014-77-C-0306
UNCLASSIFIED TR-19 NL

1 of 1

END
DATE
FILMED
10-82
DTIC

(12)

# SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL

BY

T. SELLKE and D. SIEGMUND



TECHNICAL REPORT NO. 19
JULY 1982

PREPARED UNDER CONTRACT
N00014-77-C-0306 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH



DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

# SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL

By

T. Sellke and D. Siegmund
Stanford University

TECHNICAL REPORT NO. 19

July 1982

Prepared Under the Auspices
of
Office of Naval Research Contract
N00014-77-C-0306 (NR-042-373)

DEPARTMENT OF STATISTICS
Stanford University
Stanford, California

Also issued as Technical Report No.191 under National Science Foundation Grant MCS 80-24649, Stanford University, Department of Statistics.

## SUMMARY

For the proportional hazards model of survival analysis, an appropriate large sample theory is developed for cases of staggered entry and sequential analysis. The principal techniques involve an approximation of the score process by a suitable martingale and a random rescaling of time based on the observed Fisher information. As a result we show that the maximum partial likelihood estimator behaves asymptotically like Brownian motion.





Key Words: Proportional hazards model, sequential analysis.

SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL

T. Sellke and D. Siegmund
Stanford University

## 1. Introduction

The proportional hazards model of survival analysis and its analysis by the method of partial likelihood originate in the work of Cox (1972, 1975), who argued that under general conditions maximum partial likelihood estimators have asymptotically normal distributions very similar to the asymptotic distributions of ordinary maximum likelihood estimators. Since then a number of authors have given more systematic discussions of central limit results for survival analysis. See Gill (1980), Tsiatis (1981a), and Andersen and Gill (1981).

In this paper we are concerned with related questions in the context of controlled clinical trials which possess the following two important features: (a) entry into the trial occurs at different times for different patients and (b) it seems desirable to observe the data sequentially so that the trial may be terminated at the earliest possible moment, if large treatment effects appear to be present. The authors cited above use as their starting point Cox's observation that the derivative of the log partial likelihood is a martingale, to which an appropriate martingale central limit theorem may be applied. However, with sequential observation and staggered entry, this process is no longer in general a martingale, and the approach breaks down. We shall show that it can be approximated by a martingale uniformly in time, in order to conclude that the

process of maximum partial likelihood estimates observed in a certain time scale behaves like a Brownian motion process.

Previous work on this problem seems to be limited to a Monte Carlo study by Gail, DeMets, and Slud (1981), the paper of Tsiatis (1981b), and a recent manuscript of Slud (1982). Although Slud is concerned with the special case of testing a simple null hypothesis, there is some overlap with our work, which we discuss later.

The results of this paper are not unexpected. However, it is quite surprising to find their complete justification to be so difficult, particularly in comparison to proofs of the superficially similar results of the authors cited above. In this regard it is interesting to note that Jones and Whitehead (1979) are willing to accept a cursory justification for related results, which they regard as almost obvious and propose to use as a basis for certain sequential tests. For the somewhat related Gehan test they offer a similar informal argument, but according to Slud and Wei (1982), their conclusion is incorrect in this case. Our methods do not provide satisfactory results concerning the joint distribution of the Gehan statistic, which seems to involve a substantially more difficult problem. The multivariate case is also more difficult - even in its formulation - and our results here are not yet complete.

## 2. Notation and Formulation of the Problem

Assume that $n$ patients enter a clinical trial at times $y_1, y_2, \ldots, y_n$, which may be nonrandom or occur according to an arbitrary point process. Associated with the $i^{th}$ patient is a triple $(z_i, x_i, c_i)$, where $z_i$ is a covariate, $x_i$ is the survival time following entry into the trial, and $c_i$

(possibly infinite) is a censoring variable. Thus the $i^{th}$ person is on test until time $y_i + x_i \wedge c_i$. If $x_i \leq c_i$, he dies while on the test, and the value of $x_i$ is recorded. Otherwise that observation is censored and it is known only that $x_i > c_i$. At any time $t$ there is in effect a second censoring variable $(t-y_i)^+$ in the sense that the time on test of patient $i$ prior to $t$ is $x_i \wedge c_i \wedge (t-y_i)^+$. We shall refer to $x_i$, $c_i$, and $(t-y_i)^+$ as "age" variables - the age of the $i^{th}$ patient at death, at censoring, and at time $t$ respectively.

Our basic stochastic assumptions are that $(z_i, x_i, c_i)$, $i=1, 2, \ldots, n$ are independent and identically distributed and are independent of the arrivals $y_1, \ldots, y_n$. We assume also that the $z_i$ are uniformly bounded and that given $z_i$, $x_i$ and $c_i$ are conditionally independent with $x_i$ having a cumulative hazard function of the form

$$(1) \qquad d\Lambda_i(s) = e^{\beta z_i} \lambda(s)ds$$

for some unknown parameter $\beta$ and baseline hazard function $\lambda$. For some results $z_i$ can be a vector; then $\beta$ is also, and $\beta z_i$ denotes the scalar product. For simplicity of presentation we consider explicitly only the scalar case.

All probabilities and expectations should be considered as conditional, given $y_1, y_2, \ldots, y_n$.

It is convenient to introduce the notation

$$(2) \qquad N_i(t,s) = I_{\{y_i+x_i \leq t,\ x_i \leq c_i,\ x_i \leq s\}} \qquad (s \leq t)$$

to indicate that the $i^{th}$ patient arrived and died before time $t$, that he was uncensored and was of age $\leq s$ at the time of death. We also define the set of patients at risk at time $t$ and age $s$ by

(3) $$R(t,s) = \{i: y_i \le t-s,\ x_i \wedge c_i \ge s\} \qquad (s \le t) .$$

With this notation Cox's (1975) log partial likelihood for $\beta$ can be expressed by

(4) $$\ell_n(t,\beta) = \sum_{i=1}^{n} \int_{[0,t]} \{\beta z_i - \log(\sum_{j\in R(t,s)} e^{\beta z_j})\}\, N_i(t, ds) .$$

Differentiating (4) with respect to $\beta$ gives the score process

(5) $$\dot{\ell}_n(t,\beta) = \sum_{i=1}^{n} \int_{[0,t]} \left\{ z_i - \frac{\sum_{j\in R(t,s)} z_j e^{\beta z_j}}{\sum_{j\in R(t,s)} e^{\beta z_j}} \right\} N_i(t, ds) .$$

The maximum partial likelihood estimator of $\beta$ is the solution $\hat{\beta} = \hat{\beta}_n(t)$ of

$$\dot{\ell}_n(t,\beta) = 0 .$$

Tests of the hypothesis $H_0: \beta=\beta_0$ can be based on $\hat{\beta}$ or directly on $\dot{\ell}_n(t,\beta_0)$. The usual Taylor series approximation

(6) $$0 = \dot{\ell}_n(t,\hat{\beta}) = \dot{\ell}_n(t,\beta) + (\hat{\beta}-\beta)\, \ddot{\ell}_n(t,\beta) + \ldots$$

indicates that the asymptotic behavior of $\hat{\beta}$ is intimately associated with that of $\dot{\ell}_n(t,\beta)$, which we now consider.

Let $F_{t,s}$ denote the class of events at time $t$ and age $s$, i.e. $F_{t,s}$ is the $\sigma$-algebra generated by $I_{\{y_i \le t\}}$, $y_i I_{\{y_i \le t\}}$, $z_i I_{\{y_i \le t\}}$, $I_{\{x_i \le (t-y_i)^+ \wedge s \wedge c_i\}}$, $x_i I_{\{x_i \le (t-y_i)^+ \wedge s \wedge c_i\}}$, $I_{\{c_i < (t-y_i)^+ \wedge s \wedge x_i\}}$, $c_i I_{\{c_i < (t-y_i)^+ \wedge s \wedge x_i\}}$, $i=1, 2, \ldots, n$. For each $i=1, 2, \ldots, n$, by (2), (3), and the conditional independence of $c_i$ and $x_i$, given $z_i$, as $\Delta \to 0$

$$E(N_i(t,\ s+\Delta) - N_i(t,s)|F_{t,s})$$

$$= I_{\{i\epsilon R(t,s)\}} \int_{(s,s+\Delta]} \frac{P\{c_i \geq u|z_i\}\ P\{x_i \epsilon\ du|z_i\}}{P\{c_i \geq s|z_i\}\ P\{x_i \geq s|z_i\}} + o(\Delta)\ ,$$

so at least under the additional assumption of continuity of the conditional distributions of $c_i$ given $z_i$,

(7) $$E\{N_i(t,s+\Delta) - N_i(t,s)|F_{t,s}\} = I_{\{i\epsilon R(t,s)\}}\ \{\Lambda_i(s+\Delta) - \Lambda_i(s)\} + o(\Delta)\ ,$$

where $\Lambda_i$ is given by (1). It follows from (7) that for any fixed $t$,

(8) $$\{N_i(t,s) - \Lambda_i\{(t-y_i)^+ \wedge x_i \wedge c_i \wedge s\},\ F_{t,s}\}$$

is a martingale in $s$ (cf. Gill, 1980, p. 14 or Lipster and Shiryayev, 1978, p 245).

Let $A_i(t,ds) = I_{\{i\epsilon R(t,s)\}}\ \Lambda_i(ds)$ and

(9) $$\dot{\ell}_n(t,s,\beta) = \sum_{i=1}^{n} \int_{[0,s]} \left\{ z_i - \frac{\sum_{j\epsilon R(t,u)} z_j\ e^{\beta z_j}}{\sum_{j\epsilon R(t,u)} e^{\beta z_j}} \right\} \{N_i(t,du) - A_i(t,du)\}\ .$$

It follows from (5) and simple algebra that

$$\dot{\ell}_n(t,t,\beta) = \dot{\ell}_n(t,\beta)\ .$$

Moreover, the stochastic integral in (9) inherits the martingale property of (8), so for each fixed $t$

(10) $$\{\dot{\ell}_n(t,s,\beta),\ F_{t,s}\}$$

is a martingale in $s$ (Gill, 1980, p. 10 or Lipster and Shiryayev, 1978, p. 268).

This martingale property in $s$ of $\dot{\ell}_n(t,s,\beta)$ has been the fruitful starting point for an analysis of the asymptotic normality of $\dot{\ell}_n(t,\beta) = \dot{\ell}_n(t,t,\beta)$ at one fixed point in time (cf. Gill, 1980, or Andersen and Gill, 1981). However, it does not provide useful information about the joint distribution of $\dot{\ell}_n(t,\beta)$ at different values of $t$. It is easy to show by examples that for general entry times the process $\dot{\ell}_n(t,\beta)$ is not a martingale in $t$, and hence it is necessary to uncover some additional structure before considering central limit theorems.

Let $N_i(t) = N_i(t,t)$ and $F_t = F_{t,t}$. An argument similar to that leading to (7) shows that

$$(11) \qquad E\{N_i(t+\Delta) - N_i(t) | F_t\}$$

$$= I_{\{i \in R(t,(t-y_i)^+)\}} \{\Lambda_i(t-y_i+\Delta) - \Lambda_i(t-y_i)\} + o(\Delta) .$$

Hence, with the notation $A_i(dt) = I_{\{i \in R(t,(t-y_i)^+)\}} \Lambda_i(dt-y_i)$, we see that

$$(12) \qquad \{N_i(t) - A_i(t), F_t\}$$

is a martingale in $t$.

Often stochastic integration with respect to the martingale in (12) preserves the martingale property (cf. Lipster and Shiryayev, 1978, p. 268). More precisely for our purposes we have

Lemma 1. Assume $h_i(s)$ is bounded, $F_s$-measurable, and left continuous in $s$. Then

$$\{\int_{[0,t]} h_i(s)\{N_i(ds) - A_i(ds)\},\ F_t\}$$

is a martingale in $t$.

By a change of variable in (9)

$$(13) \qquad \dot{\ell}_n(t,\beta) = \sum_{i=1}^{n} \int_{[0,t]} \left\{ z_i - \frac{\sum_{j\in R(t,u-y_i)} z_j e^{\beta z_j}}{\sum_{j\in R(t,u-y_i)} e^{\beta z_j}} \right\} \{N_i(du) - A_i(du)\} .$$

Although (13) is a stochastic integral, Lemma 1 does not apply because the integrands are not $F_u$-measurable and they depend on $t$. However, an informal law of large numbers argument suggests that these integrands are approximately

$$\left\{ z_i - \frac{E(z_1 e^{\beta z_1};\ x_1 \wedge c_1 \geq u-y_i)}{E(e^{\beta z_1};\ x_1 \wedge c_1 \geq u-y_i)} \right\} I_{\{y_i<u\}} ,$$

which does not depend on $t$; and according to Lemma 1

$$(14) \qquad \sum_{i=1}^{n} \int_{[0,t]} \left\{ z_i - \frac{E(z_1 e^{\beta z_1};\ x_1 \wedge c_1 \geq u-y_i)}{E(e^{\beta z_1};\ x_1 \wedge c_1 \geq u-y_i)} \right\} \{N_i(du) - A_i(du)\}$$

$$= \sum_{i=1}^{n} \int_{[0,t]} \left\{ z_i - \frac{E(z_1 e^{\beta z_1};\ x_1 \wedge c_1 \geq s)}{E(e^{\beta z_1};\ x_1 \wedge c_1 \geq s)} \right\} \{N_i(t,ds) - A_i(t,ds)\}$$

is a martingale in $t$.

In broad outline the goals of the rest of this paper are to show that the martingale in (14) is a good approximation to the score process (13) uniformly in $t$ as $n\to\infty$ (Theorem 1), and to apply a martingale central limit theorem to show that (14) (and hence (13)) suitably rescaled behave

like a Brownian motion process asymptotically as $n \to \infty$ (Theorem 2). Finally the asymptotic behavior of $\hat{\beta}$ is related to that of $\dot{\ell}_n(t,\beta)$ via (6) and a consistency argument (Theorem 3).

Before proceeding with the technical developments to follow, we make some remarks related to the asymptotic rescaling of the martingale (14). In order that (14) behave asymptotically like a Brownian motion process it is important that its "quadratic variation" or its "predictable quadratic variation" grow approximately linearly in $t$. (See Meyer, 1976, p. 267 for the definition of these terms in general and (33) for the special case of (14).) For (14) this linear growth does not occur in the time scale $t$, and it is convenient to introduce a data dependent transformation of time to obtain the desired linearity. From a statistical point of view the natural mechanism to effect this change of time is the observed Fisher information or minus the second derivative of the log (partial) likelihood, $-\ddot{\ell}_n(t,\beta)$, which will be shown to be essentially the same as the quadratic variation of (14). Hence, for $v > 0$ let

$$\tau_n(v,\beta) = \inf\{t: -\ddot{\ell}_n(t,\beta) \geq vn\} \ . \tag{15}$$

Theorem 2 of Section 4 asserts that

$$n^{-\frac{1}{2}} \dot{\ell}_n(\tau_n(v,\beta), \beta) \xrightarrow{\mathcal{L}} W(v) \ , \tag{16}$$

where $W$ is a standard Brownian motion. Of course, in practice one must actually use the observable quantity $\tau_n(v,\hat{\beta})$ to define the time scale. See Grambsch (1982) or Lai and Siegmund (1982) for discussions of the use of Fisher information as a means of rescaling time.

It is worth noting that much of the preceding discussion generalizes immediately to several dimensions. However, the rescaling of time indicated

by (15) and (16) carries over directly only if all elements of the matrix $-\ddot{\ell}_n(t,\beta)$ have the same growth rates in $t$ for large $n$. Except when $\beta=0$ this is typically not the case.

We now turn to a detailed analysis of the approximation of (5) by (14). A more thorough discussion of (16) is contained in Section 4.

## 3. Approximation of $\dot{\ell}_n(t,\beta)$ by the Martingale (14).

Theorem 1. Let $R_n(t)$ denote the difference between the martingale in (14) and $\dot{\ell}_n(t,\beta)$ given by (5) or equivalently by (9) with $s=t$. Then for arbitrary $\varepsilon > 0$, as $n\to\infty$

$$P\{\sup_{0<t<\infty} |R_n(t)| > \varepsilon n^{\frac{1}{2}}\} \to 0 .$$

The proof of Theorem 1 is a consequence of the following three lemmas. It will be convenient to use the notation:

$$\tilde{\mu}(t,s) = \frac{\sum_{j\in R(t,s)} z_j e^{\beta z_j}}{\sum_{j\in R(t,s)} e^{\beta z_j}} \qquad (s \le t) \tag{17}$$

and

$$\mu(s) = \frac{E(z_1 e^{\beta z_1};\ x_1 \wedge c_1 \ge s)}{E(e^{\beta z_1};\ x_1 \wedge c_1 \ge s)} . \tag{18}$$

In (17) and (18) we interpret 0/0 as 0. With this notation

$$R_n(t) = \sum_i \int_{[0,t]} \{\tilde{\mu}(t,s)-\mu(s)\}\{N_i(t,ds) - A_i(t,ds)\} . \tag{19}$$

Lemma 2. Uniformly in $0 < t < \infty$

$$E\, R_n^2(t) = O(\log n)\ .$$

Proof. From fundamental properties of stochastic integrals

$$E\, R_n^2(t) = E\Big(\sum_i \int_{[0,t]} \{\tilde{\mu}(t,s)-\mu(s)\}^2\, N_i(t,ds)\Big)\ .$$

By considering the $i^{th}$ term and conditioning on $x_i$, $R(t,x_i)$, and the event $\{x_i \leq c_i \wedge (t-y_i)^+\}$, we obtain

$$(20)\quad E\, R_n^2(t) = E\Big(\sum_i N_i(t,x_i)\, E[\{\tilde{\mu}(t,x_i)-\mu(x_i)\}^2 | x_i,\ R(t,x_i),\ x_i \leq (t-y_i)^+ \wedge c_i]\Big)$$

$$\leq \text{const. } E\Big(\sum_i \frac{N(t,x_i)}{|R(t,x_i)|^2}\, E[\{\mu_0(x_i) \sum_{j\in R(t,x_i)} z_j e^{\beta z_j}$$

$$- \mu_1(x_i) \sum_{j\in R(t,x_i)} e^{\beta z_j}\}^2 | x_i,\ R(t,x_i),\ x_i \leq c_i \wedge (t-y_i)]\Big)\ ,$$

where $\mu_\nu(s) = E(z_1^\nu e^{\beta z_1} | x_1 \wedge c_1 \geq s)$ for $\nu=0$ and $1$, and $|A|$ denotes the cardinality of the set $A$. Let $R_i^*(t,s) = R(t,s) - \{i\}$, and observe that given $x_i$, $x_i \leq c_i \wedge (t-y_i)^+$, and $R_i^*(t,x_i) = \{j_1, \ldots, j_r\}$, $z_{j_1}, \ldots, z_{j_r}$ are independent and identically distributed with

$$E(z_{j_\ell}^\nu e^{\beta z_{j_\ell}} | x_i,\ R_i^*(t,x_i),\ x_i \leq c_i \wedge (t-y_i)^+) = \mu_\nu(x_i)$$

for $\nu=0$ or $1$. Hence except for the terms involving $i$, the conditional expectations on the right hand side of (20) involve the square of a sum of i.i.d. random variables having mean $0$. Hence

$$E\,R_n^2(t) \leq \text{const.}\{E(\sum_i \frac{N_i(t,x_i)}{|R(t,x_i)|}) + E(\sum_i \frac{N_i(t,x_i)}{|R(t,x_i)|^2})\}$$

$$= E[0\{\log \sum_i N_i(t,t)\}] = 0(\log n)$$

uniformly in $t$.

Lemma 3. Let $0 < \varepsilon < 1/10$ and $0 = t_0 < t_1 < \ldots < t_{n^{1-3\varepsilon}} = \infty$. Then $P\{\max_{1\leq k\leq n^{1-3\varepsilon}} |R_n(t_k)| > n^{\frac{1}{2}(1-\varepsilon)}\} = 0(n^{-2\varepsilon} \log n)$.

The proof of Lemma 3 is an immediate consequence of Lemma 2 and Chebyshev's inequality. It remains to make a specific choice of the points $\{t_k\}$ and show that $R_n(t)$ cannot change too rapidly between these points.

Let $D_k = \sum_i \int_{[t_{k-1},t_k)} N_i(t_k,ds-y_i)$ denote the number of deaths observed during $[t_{k-1},t_k)$ and let

$$H_k = \sum_i \int_{[t_{k-1},t_k)} A_i(t_k,\, ds-y_i)$$

denote the associated accumulated hazard. We choose the partition $\{t_k,\ k=0, 1, \ldots, n^{1-3\varepsilon}\}$ so that for all $k$

$$E\,D_k \leq 2n^{3\varepsilon}\,. \tag{21}$$

Lemma 3 above and Lemma 4 below complete the proof of Theorem 1.

Lemma 4. $P\{\max_k \max_{t_{k-1}\leq t<t_k} |R_n(t)-R_n(t_{k-1})| > n^{\frac{1}{2}(1-\varepsilon)}\} \to 0$ as $n\to\infty$.

Proof. Note that if $\mathcal{B}(p)$ is a Bernoulli variable with parameter $p$, then for $p \leq \frac{1}{2}$, $\mathcal{B}(p)$ is stochastically smaller than a Poisson variable with parameter $2p$, say $\mathcal{P}(2p)$. Hence for all $0 \leq p \leq 1$, $\mathcal{B}(p)$ is stochastically smaller than $2p + \mathcal{P}(2p)$. Since $D_k$ is a sum of independent

Bernoulli variables, it follows from (21) that $D_k$ is stochastically smaller than $4n^{3\epsilon} + \mathcal{P}(4n^{3\epsilon})$. Hence by easy large deviation estimates

$$(22) \qquad P\{\max_k D_k > n^{7\epsilon/2}\} \to 0 \qquad (n\to\infty) .$$

On $\{H_k \geq n^{4\epsilon}\}$ $D_k$ is stochastically larger than a Poisson random variable of mean $n^{4\epsilon}$ and hence $P\{D_k \leq n^{7\epsilon/2}, H_k \geq n^{4\epsilon}\} \leq P\{\mathcal{P}(n^{4\epsilon}) \leq n^{7\epsilon/2}\} = O(n^{-1})$. From this and (22) it follows that

$$(23) \qquad P\{\max_k H_k > n^{4\epsilon}\} \to 0 \qquad (n\to\infty) .$$

Let $t_{k-1} \leq t < t_k$. By (19)

$$(24) \quad R_n(t)-R_n(t_{k-1}) = \int_{[t_{k-1},t]} \{\tilde{\mu}(t,s)-\mu(s)\}\{N(t,ds)-A(t,ds)\}$$

$$+ \int_{[0,t_{k-1})} \{\tilde{\mu}(t,s)-\mu(s)\}\{N(t,ds)-A(t,ds)-N(t_{k-1},ds)+A(t_{k-1},ds)\}$$

$$+ \int_{[0,t_{k-1})} \{\tilde{\mu}(t,s)-\tilde{\mu}(t_{k-1},s)\}\{N(t_{k-1},ds)-A(t_{k-1},ds)\} ,$$

where we have used the notation

$$N(t,ds) = \sum_i N_i(t,ds), \ A(t,ds) = \sum_i A_i(t,ds) .$$

By assumption the $z_i$ are bounded and hence $\tilde{\mu}(t,s)$ and $\mu(s)$ are bounded. Since $N(t,ds) - N(t_{k-1},ds)$ and $A(t,ds) - A(t_{k-1},ds)$ are both positive and increasing in $t$, it follows that the first two terms on the right hand side of (24) are of order $D_k+H_k$ uniformly in $t_{k-1} \leq t < t_k$. Hence by (22) and (23) it suffices to consider the third integral in (24). Let

(25) $$m(t,s) = \sum_{j \in R(t,s)} e^{\beta z_j}$$

and observe that

(26) $$A(t,ds) = m(t,s)\ \lambda(s)ds\ .$$

Letting $B/2$ denote a bound on the $z_i$, we find from (17) and some algebra that uniformly in $t_{k-1} \leq t < t_k$

(27) $$|\tilde{\mu}(t,s)-\tilde{\mu}(t_{k-1},s)| \leq B\{m(t_k,s)-m(t_{k-1},s)\}/m(t_k,s)\ .$$

Hence by (24) and (27) it suffices to show

(28) $$P\{\max_k \int_{[0,t_{k-1})} [\{m(t_k,s)-m(t_{k-1},s)\}/m(t_k,s)]\ N(t_{k-1},ds) > n^{\frac{1}{2}(1-\varepsilon)}\} \to 0$$

and

(29) $$P\{\max_k \int_{[0,t_{k-1})} [\{m(t_k,s)-m(t_{k-1},s)\}/m(t_k,s)]\ A(t_{k-1},ds) > n^{\frac{1}{2}(1-\varepsilon)}\} \to 0\ .$$

From (26) and some algebra we see that the random variable in (29) is majorized by $\max_k H_k$, so (29) follows from (23).

Now consider (28). It is easily seen by direct calculation that

$$L_k(s) = \int_{[0,s)} [\{m(t_k,u) - m(t_{k-1},u)\}/m(t_k,u)]\ N(t_{k-1},du)$$
$$- \{N(t_k,s) - N(t_{k-1},s)\}$$

is a supermartingale for $0 \leq s \leq t_{k-1}$, which changes by jumps downward of size 1 and upward of size at most equal to 1. Furthermore, $N(t_k,t_{k-1}) - N(t_{k-1},t_{k-1}) \leq D_k$, so by (22), to prove (28) it suffices to show

(30) $$P\{\max_k L_k(t_{k-1}) > n^{4\epsilon}\} \to 0 \ .$$

Let $S_o = 0$, and for $j=1, 2, \ldots$ let

$$S_j = \inf\{s\colon s \geq S_{j-1},\ L_k(s) - L_k(S_{j-1}) \geq 1 \text{ or } < 0\}$$

where it is understood that $\inf \phi = t_{k-1}$. Obviously $-1 \leq L_k(S_j) - L_k(S_{j-1}) \leq 2$, and from the supermartingale property we see that on $\{S_{j-1} < t_{k-1}\}$

$$E\{L_k(S_j) - L_k(S_{j-1}) | F_{t_k, S_{j-1}}\} \leq 0 \ ,$$

and hence

(31) $$P\{S_j < t_{k-1},\ L_k(S_j) - L_k(S_{j-1}) \geq 1 | F_{t_k, S_{j-1}}\} \leq 1/2 \ .$$

It follows from (31) that between downward jumps the total increase of $L_k(s)$ is stochastically smaller than $1+y$, where $P\{y=m\} = (1/2)^{m+1}$, $m=0, 1, \ldots$ . Since the total number of downward jumps is $D_k$, an easy large deviation estimate gives

$$P\{L_k(t_{k-1}) \geq n^{4\epsilon},\ D_k \leq n^{7\epsilon/2}\} = o(n^{-1}) \ .$$

Hence by (22)

$$P\{\max_k L_k(t_{k-1}) > n^{4\epsilon}\} \leq P\{\max_k D_k > n^{7\epsilon/2}\} + \sum_k P\{L_k(t_{k-1}) \geq n^{4\epsilon},\ D_k < n^{7\epsilon/2}\}$$

$\to 0$ as $n\to\infty$.

## 4. Convergence to Brownian Motion

In this section we give a precise interpretation of (16) and indicate its proof. Let $Q_n(t,\beta)$ denote the martingale in (14), which in the notation of Section 3 can be written

(32) $$Q_n(t,\beta) = \sum_i \int_{[0,t]} \{z_i - \mu(s-y_i)\} \{N_i(ds) - A_i(ds)\}$$
$$= \sum_i \int_{[0,t]} \{z_i - \mu(s)\} \{N_i(t,ds) - A_i(t,ds)\} .$$

It follows from the first representation of $Q_n$ in (32) and the independence of the different terms that the predictable quadratic variation of the Martingale $Q_n$ is

(33) $$\sum_i \int_{[0,t]} \{z_i-\mu(s-y_i)\}^2 A_i(ds) = \sum_i \int_{[0,t]} \{z_i-\mu(s)\}^2 A_i(t,ds) .$$

Let $v_f = E[e^{\beta z_1}\int_0^{x_1 \wedge c_1}\{z_1-\mu(s)\}^2\lambda(s)ds]$ and note that by the law of large numbers

$$n^{-1} \sum_i \int_{[0,\infty)} \{z_i-\mu(s)\}^2 A_i(\infty,ds) \to v_f$$

in probability. Hence for $0 < v < v_f$ and $T_n(v,\beta)$ defined by

$$T_n(v,\beta) = \inf\{t: n^{-1} \sum_i \int_{[0,t]} \{z_i-\mu(s)\}^2 A_i(t,ds) \geq v\}$$

we have $P\{T_n(v,\beta) < \infty\} \to 1$ and

(34) $$n^{-1}\sum_i \int_{[0,T_n(v,\beta)]} \{z_i-\mu(s)\}^2 A_i(t,ds) \to v$$

in probability. It follows from (34) and the form of the martingale central limit theorem given by Rebolledo (1980, p. 273, Proposition 1) that for every $0 < v < v_f$

(35) $$n^{-\frac{1}{2}} Q_n(T_n(\cdot,\beta), \beta) \xrightarrow{\mathcal{D}} W(\cdot)$$

on $[0,v]$ as $n\to\infty$, where $W$ is a standard Brownian motion.

This result is unsatisfactory for statistical purposes because (33) is not an observable random variable - even under a simple null hypothesis, when $\beta$ is assumed to be known. Consider now

$$(36) \qquad -\ddot{\ell}_n(t,\beta) = \int_{[0,t]} \tilde{\sigma}^2(t,s)\, N(t,ds) \ ,$$

where

$$\tilde{\sigma}^2(t,s) = \sum_{j\epsilon R(t,s)} (z_j - \tilde{\mu}(t,s))^2\, e^{\beta z_j} / \sum_{j\epsilon R(t,s)} e^{\beta z_j} \ ,$$

and let $\tau_n(v,\beta)$ be defined by (15).

<u>Theorem 2</u>. For each $0 < v < v_f$,

$$P\{\tau_n(v,\beta) < \infty\} \to 1$$

and on $[0,v]$

$$(37) \qquad n^{-\frac{1}{2}}\, \dot{\ell}_n(\tau_n(\cdot,\beta),\, \beta) \xrightarrow{\mathcal{D}} W(\cdot) \ ,$$

where $W(\cdot)$ is standard Brownian motion.

The key tools in the proof of Theorem 2 are Theorem 1, which shows that it suffices to prove (37) with $Q_n$ in place of $\dot{\ell}_n$, and Lemma 5 below, which shows that (34)holds with $\dot{\tau}_n$ in place of $T_n$, so the martingale central limit theorem applies to yield (35) with $\tau_n$ in place of $T_n$.

<u>Lemma 5</u>. For small positive $\epsilon$, as $n \to \infty$

$$(38) \qquad P\{\max_t |\ddot{\ell}_n(t,\beta) + \sum_i \int_{[0,t]} \{z_i - \mu(s)\}^2\, A_i(t,ds)| > n^{1-\epsilon}\} \to 0 \ .$$

<u>Proof</u>. The proof is similar to Theorem 1, so we give only a general outline. Observe that

$$(39)\qquad \ddot{\ell}_n(t,\beta) + \sum_i \int_{[0,t]} \{z_i-\mu(s)\}^2 A_i(t,ds)$$

$$= \int_{[0,t]} \{\tilde{\mu}(t,s)-\mu(s)\}^2 A(t,ds) - \int_{[0,t]} \sigma^2(s)\{N(t,ds)-A(t,ds)\}$$

$$- \int_{[0,t]} \{\tilde{\sigma}^2(t,s)-\sigma^2(s)\}\{N(t,ds)-A(t,ds)\} \quad ,$$

where

$$\sigma^2(s) = E[\{z_1-\mu(s)\}^2 e^{\beta z_1};\ x_1 \wedge c_1 \geq s]/E(e^{\beta z_1};\ x_1 \wedge c_1 \geq s) \quad .$$

Each of the terms on the right hand side of (39) can be estimated by techniques similar to the proof of Theorem 1. For example, the third integral can be split into a part involving a difference of second moments and a part involving a difference of squares of first moments. The second moment piece is treated directly by the techniques of Theorem 1; for the first moment piece we use $a^2 - b^2 = (a+b)(a-b)$ and the boundedness of $(a+b)$ in order to apply the techniques of Theorem 1.

## 5. Discussion

In order to turn Theorem 2 into a statistically interesting result, one must (a) relate the behavior of the partial likelihood function given in (37) to that of the maximum partial likelihood estimator $\hat{\beta}$ defined by $\dot{\ell}_n(t,\hat{\beta}) = 0$, and (b) replace $\beta$ by $\hat{\beta}$ in (15), so that the desired renormalization of time is accomplished by observable random variables. This yields the main result of the paper.

<u>Theorem 3</u>. Let $0 < v < v_f$ and

$$\hat{\tau}_n(v) = \inf[t: -\ddot{\ell}_n\{t, \hat{\beta}_n(t)\} \geq vn] .$$

For $0 < v_* < v^* < v_f$ as $n \to \infty$

$$(\cdot)n^{\frac{1}{2}}[\hat{\beta}_n\{\hat{\tau}_n(\cdot) - \beta] \overset{\mathcal{L}}{\to} W(\cdot) \tag{40}$$

or equivalently

$$-n^{-\frac{1}{2}}\ddot{\ell}_n[\hat{\tau}_n(\cdot), \hat{\beta}_n\{\hat{\tau}_n(\cdot)\}] [\hat{\beta}_n\{\hat{\tau}_n(\cdot)\} - \beta] \to W(\cdot) \tag{41}$$

on $[v_*, v^*]$.

We have omitted the proof of this theorem. Basically one combines the results of Sections 3 and 4 with some Taylor series expansions to prove the consistency of $\hat{\beta}$, and then uses this consistency, (6), and Theorem 2 to prove Theorem 3.

A consequence of Theorem 3 is that in the time scale determined by $\hat{\tau}_n$, $\hat{\beta}_n$ can be approximated by a Brownian motion process for the purpose of sequentially testing statistical hypotheses about $\beta$ or for estimating $\beta$. Thus if we would be satisfied with a repeated significance test or a truncated sequential probability ratio test for testing a hypothesis about the drift of a Brownian motion process, we can obtain an analogous asymptotic test for $\beta$.

Since $n^{\frac{1}{2}} W(v)$ and $W(nv)$, $0 \leq v \leq \infty$ have the same joint distributions, it is tempting to rewrite (40) as

$$nv[\hat{\beta}_n\{\hat{\tau}_n(v)\} - \beta] \sim W(nv) \quad \text{for} \quad nv_* < nv < nv^* \tag{42}$$

and treat $nv = u$ as one variable. A loose interpretation of (42) would be that

(43) $$-\ddot{\ell}_n\{t, \hat{\beta}_n(t)\}\{\hat{\beta}_n(t) - \beta\}$$

behaves approximately like

(44) $$W[-\ddot{\ell}_n\{t, \hat{\beta}_n(t)\}]$$

provided $-\ddot{\ell}_n = -\ddot{\ell}_n\{t, \hat{\beta}_n(t)\}$ is "large". Of course, the theorem specifies that "large" means proportional to $n$ with constants of proportionality bounded away from $0$ and from $v_f$.

In practice it is probably unnecessary to interpret the minimum information requirement stringently. In fact, close scrutiny of the proof of Theorem 2 shows that $n$ in (15) could be replaced by $n^{1-\delta}$ for suitable small positive $\delta$ and then the normalizing $n^{-\frac{1}{2}}$ in (37) would become $n^{-\frac{1}{2}(1-\delta)}$. Hence the approximation of (43) by (44) is valid for values of $-\ddot{\ell}$ of smaller order of magnitude than $n$, but we do not know how much smaller. We conjecture that with a proper reformulation it is possible to give an interpretation to the approximation of (43) by (44) provided only that $-\ddot{\ell}$ is large.

The maximum information requirement that $-\ddot{\ell}_n << nv_f$ is probably more important and could conceivably cause some difficulty in practice, since $v_f$ is essentially never known. However, if the patient arrival rate is sufficiently great and the experimental period comparatively short so that some reasonable percentage of the total number put on test is still alive at the end of the experiment, no problems should arise.

It seems desirable to conduct a Monte Carlo experiment to get some feeling for the practical limitations of Theorem 3. For the related problem of testing the null hypothesis $\beta=0$, Gail, DeMets, and Slud (1981)

conclude that the score statistic under the null hypothesis is reasonably approximated by a Brownian motion. Their time renormalization is not appropriate for general $\beta$, however.

Slud's (1982) theoretical approach is superficially similar to ours in that he introduces a martingale to approximate the score process of the partial likelihood. His martingale is different from ours (although it is a special case of the class of martingales described by Lemma 1). He considers only the null hypothesis $\beta=0$ and uses a time renormalization which would be inappropriate for general $\beta$. Also what corresponds to our Lemma 4 is essentially his assumption A.5. This assumption is never actually verified although Slud states that it can be verified under various sets of conditions, all of which require strong hypotheses on the arrival process.

It is not obvious how one should generalize these results to multidimensional $\beta$. Except when $\beta=0$, one cannot expect that the information about the various coordinates of $\beta$ accumulate at the same rate, and hence one cannot generalize (15) directly. In the case where one coordinate of $\beta$ is a treatment indicator, it seems possible to study this one coordinate sequentially by making a time change in terms of the residual variance of its regression on the other coordinates. We hope to discuss this problem in a future publication.

## REFERENCES


Anderson, P. K. and Gill, R. (1981). Cox's regression model for counting processes: a large sample study, submitted to Ann. Statist.

Cox, D. R. (1972). Regression models and life tables (with discussion), J. Roy. Statist. Soc. B. **34**, 187-220.

Cox, D. R. (1975). Partial likelihood, Biometrika **62**, 269-276.

Gail, M., DeMets, D., and Slud, E. (1981). Simulation studies on increments of the two-sample log rank test for survival data with application to group sequential boundaries, Proceedings of Special Topic IMS Meeting on Survival Statistics, Columbus, Ohio, October, 1981.

Gill, R. D. (1980). Censoring and Stochastic Integrals, Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

Grambsch, P. (1982). Sequential sampling based on observed Fisher information to guarantee the accuracy of the maximum likelihood estimator, submitted to Ann. Statist.

Jones, D. and Whitehead, J. (1979). Sequential forms of the log rank and modified Wilcoxon test for censored data, Biometrika **66**, 105-113.

Lai, T. L. and Siegmund, D. (1982). Fixed accuracy estimation of an autoregressive parameter, submitted to Ann. Statist.

Lipster, R. S. and Shiryayev, A. N. (1978). Statistics of Random Processes II, Springer-Verlag, New York-Heidelberg-Berlin.

Meyer, P. A. (1976). Uncours sur l'intégrales stochastique, Seminaire de Probabilités X. Universite de Strasbourg. Lect. Notes in Math 511, Springer, New York-Heidelberg-Berlin.

Rebolledo, R. (1980). Central limit theorem for local martingales, Zeit. f. Wahr. u. verw. Geb. **51**, 269-286.

Slud, E. (1982). Sequential linear rank tests for two-sample censored survival data, submitted to Ann. Statist.

Slud, E. and Wei, L. J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic, Jour. Amer. Statist. Assoc. **77**.

Tsiatis, A. (1981a). A large sample study of Cox's regression model, Ann. Statist. **9**, 93-108.

Tsiatis, A. (1981a). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time, Biometrika **68**, 311-315.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | |
|---|---|
| 1. REPORT NUMBER: 19 | 2. GOVT ACCESSION NO.: A119375 |
| 3. RECIPIENT'S CATALOG NUMBER | |
| 4. TITLE (and Subtitle): SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL | 5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT |
| | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s): T. SELLKE AND D. SIEGMUND | 8. CONTRACT OR GRANT NUMBER(s): N00014-77-C-0306 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, STANFORD, CALIF. | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042-373 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: Statistics and Probability Program Code (411(SP)), Office of Naval Research, Arlington, Va. 22217 | 12. REPORT DATE: JULY 1982 |
| | 13. NUMBER OF PAGES: 22 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | 15. SECURITY CLASS. (of this report): UNCLASSIFIED |
| | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also issued as Technical Report No. 191 under National Science Foundation Grant MCS 80-24649, Stanford University, Department of Statistics.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

PROPORTIONAL HAZARDS MODEL, SEQUENTIAL ANALYSIS.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE

REPORT #19

SUMMARY

For the proportional hazards model of survival analysis, an appropriate large sample theory is developed for cases of staggered entry and sequential analysis. The principal techniques involve an approximation of the score process by a suitable martingale and a random rescaling of time based on the observed Fisher information. As a result we show that the maximum partial likelihood estimator behaves asymptotically like Brownian motion.

DA

FIL

0—